“En lagringsinfrastrukturs behov”

Dejan Vitlacil
SNIC Storage Coordinator
KTH Royal Institute of Technology
vitlacil@kth.se

Swedish National Infrastructure for Computing

About

The Swedish National Infrastructure for Computing (SNIC) is a national research infrastructure that provides a **balanced** and **cost-efficient** set of resources resources and user support for **large scale computation** and **data storage**.

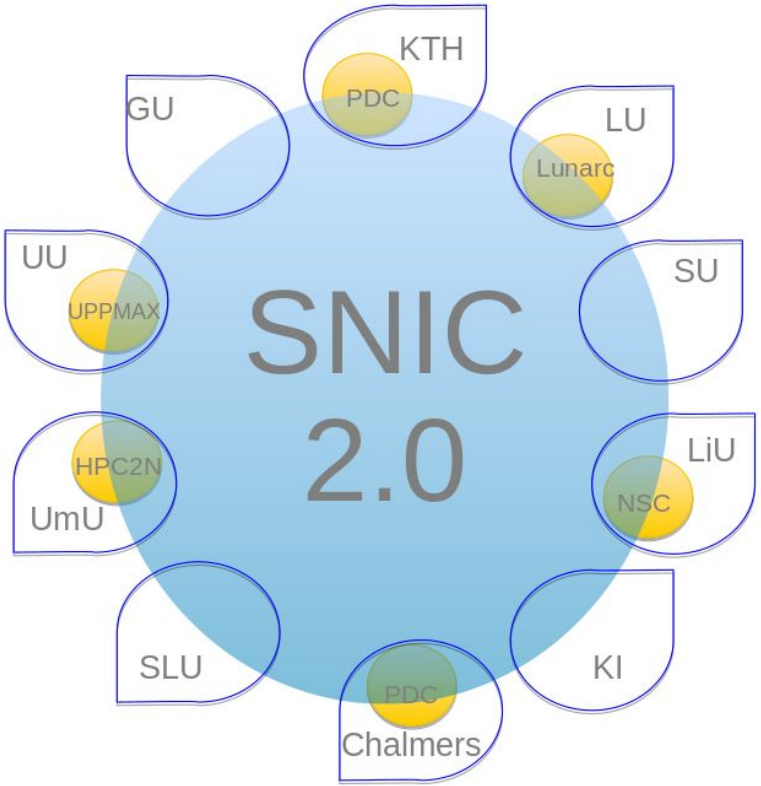
The infrastructure shall meet the needs of researchers from **many scientific disciplines** and from many institutes for higher education and many research institutes.

The resources provided by SNIC are made available through **open procedures** such that the best Swedish research is supported and new research is facilitated.

SNAC - Swedish National Allocations Committee - allocates SNIC resources.

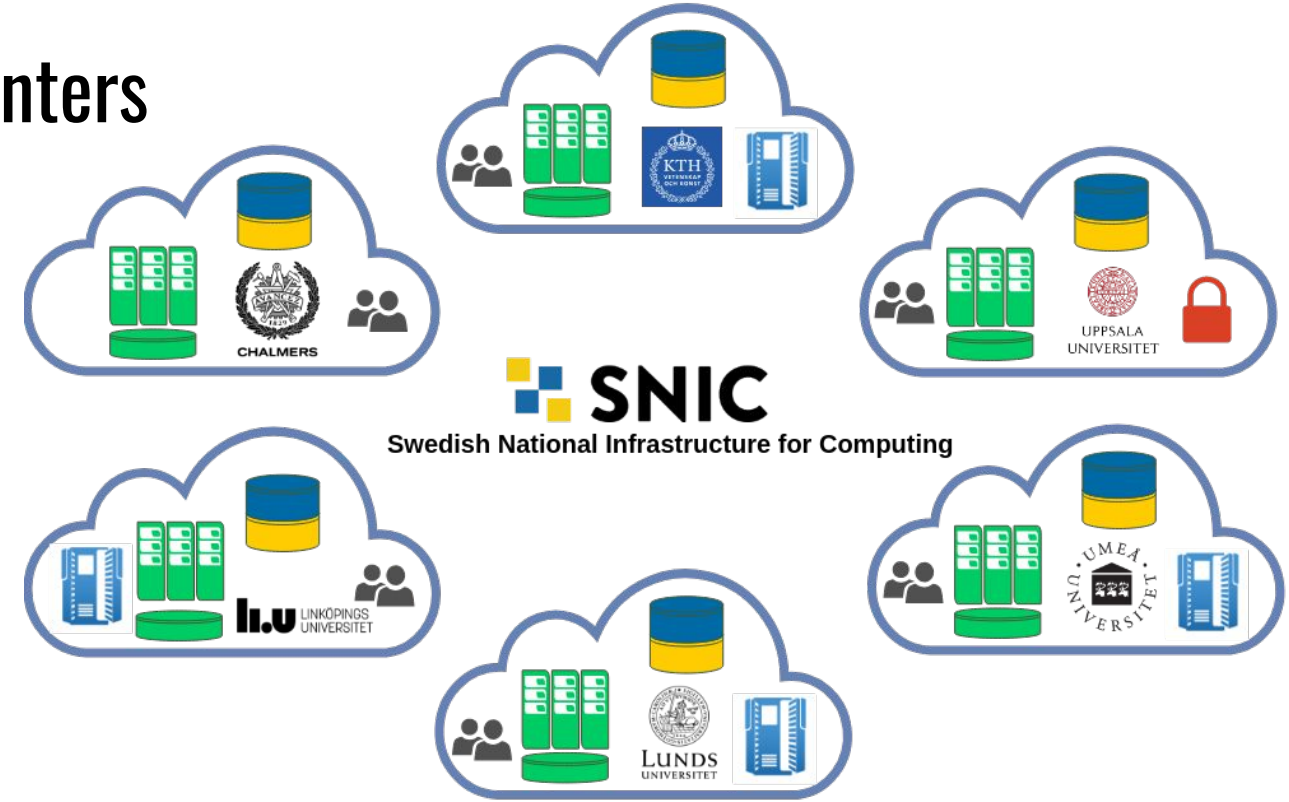
SNIC

SNIC Consortium



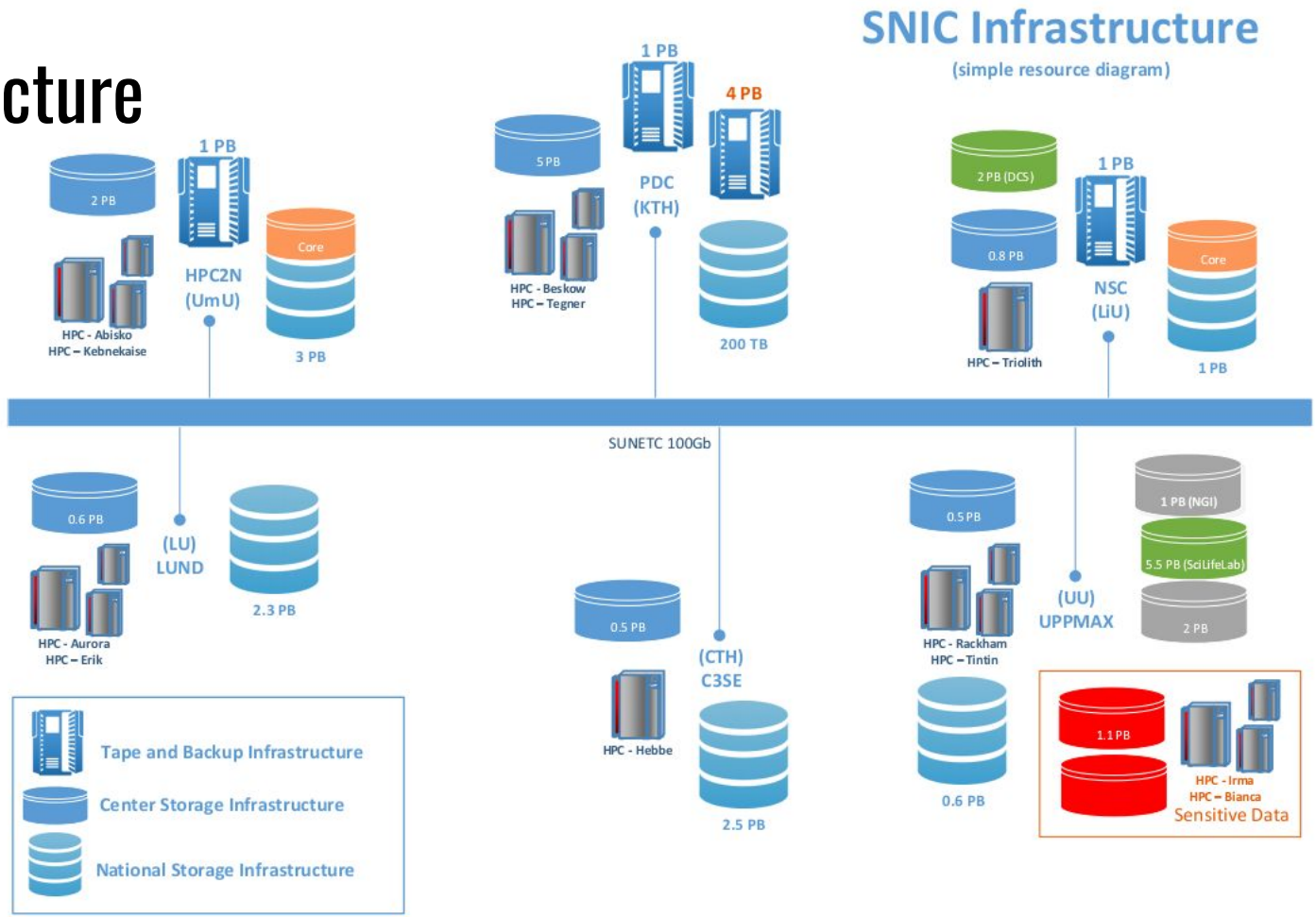
SNIC

Data Centers

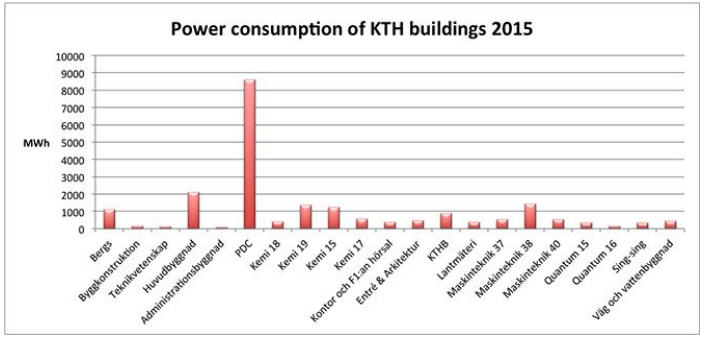


SNIC

Infrastructure

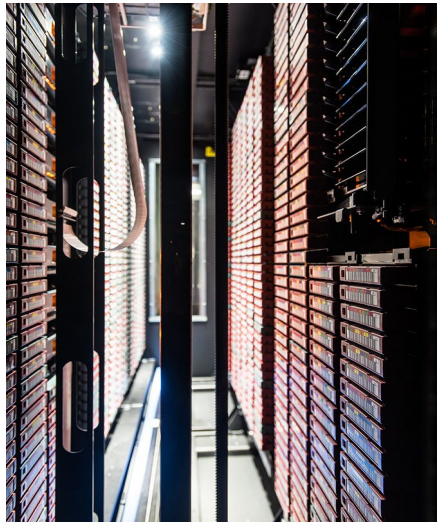


SNIC

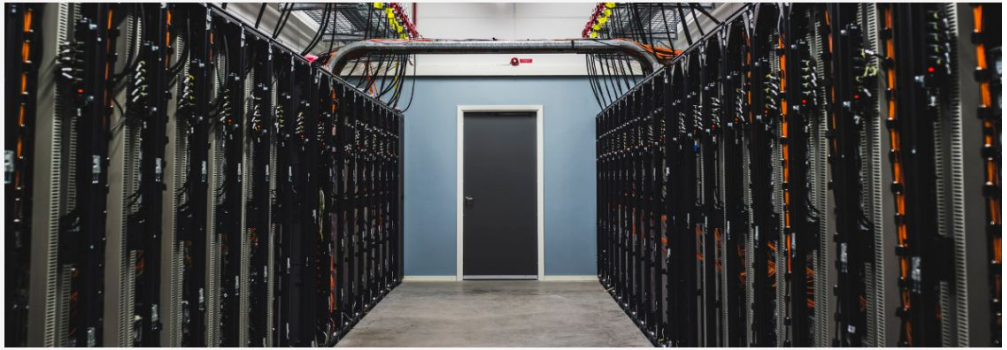

Power consumption of KTH buildings 2015
MWh
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
Berga
Byggkonstruktion
Teknikvetenskap
Huvudbyggnad
Administrationsbyggnad
PDC
Kemi 18
Kemi 19
Kemi 15
Kemi 17
Kontor och F1:an hörsal
Entré & Arkitektur
KTHB
Lantmäteri
Maskinteknik 37
Maskinteknik 38
Maskinteknik 40
Quantum 15
Quantum 16
Sing-sing
Väg och vattenbyggnad

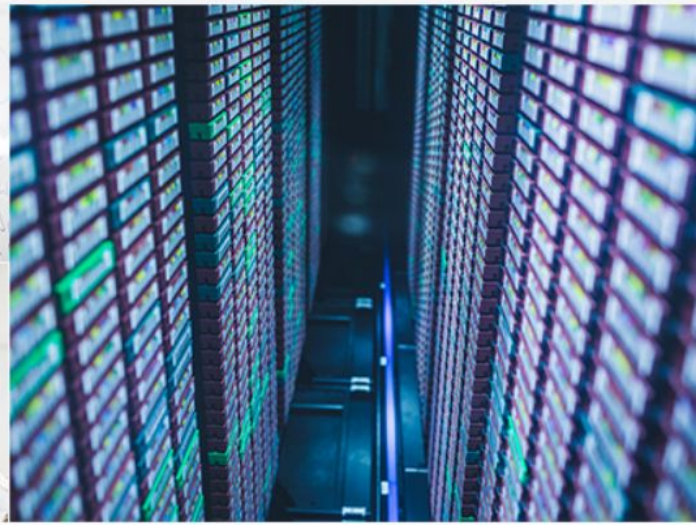

CRAY
KTH


CRAY
CRAY
CRAY
CRAY
CRAY
CRAY
KTH

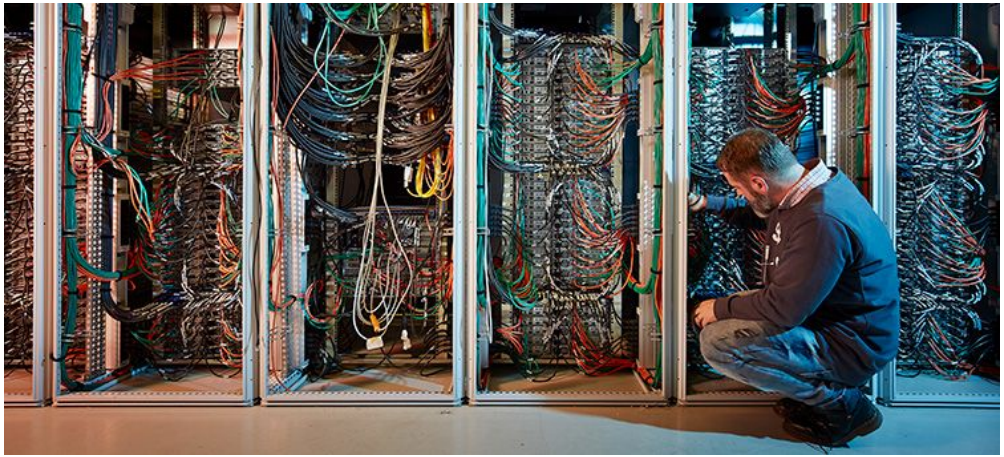


SNIC



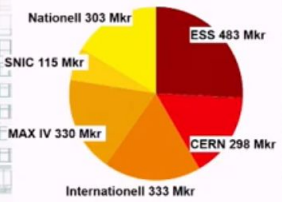


SNIC



FINANSIERING AV FORSKNINGSINFRASTRUKTUR

Vetenskapsrådet



RFIs budget 2018: 1 862 Mkr

Fördelas på:

European Spallation Source: 483 Mkr

Nationella infrastrukturer
(inkl. MAX IV och SNIC): 778 Mkr

Internationella infrastrukturer
(inkl. CERN): 631 Mkr

SNIC

Process

- Check for the open calls

SNIC SUPR

Start
Rounds
Support
Login

Your are not logged in.

Start / Rounds

Rounds

Open for Proposals

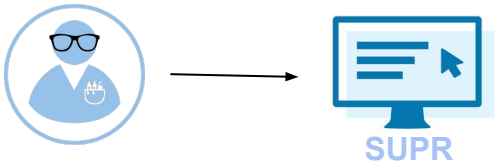
SNIC Rounds	Deadline
SNAC Large, Spring 2019	2019-04-05 15:00
SNAC Medium, 2019	—
SNAC Small C3SE, 2019	—
SNAC Small HPC2N, 2019	—
SNAC Small Lunarc, 2019	—
SNAC Small NSC, 2019	—
SNAC Small UPPMAX, 2019	—
SNAC Science Cloud 2019	—
SNAC Medium Swestore 2019	—
SNAC Small Swestore 2019	—
SIMPLER 2019	—
SNIC SENS Medium 2019	—
SNIC SENS Small 2019	—

Local Rounds	Deadline
Cross-Border Resource sharing test instance	—
LiU Local, 2019	—
L-SENS, 2019	—
LU Local, 2019	—
LVIS, 2019	—
PDC Business 2019	—
PDC Local 2019	—
UPPMAX courses 2019	—
UPPMAX Storage 2019	—

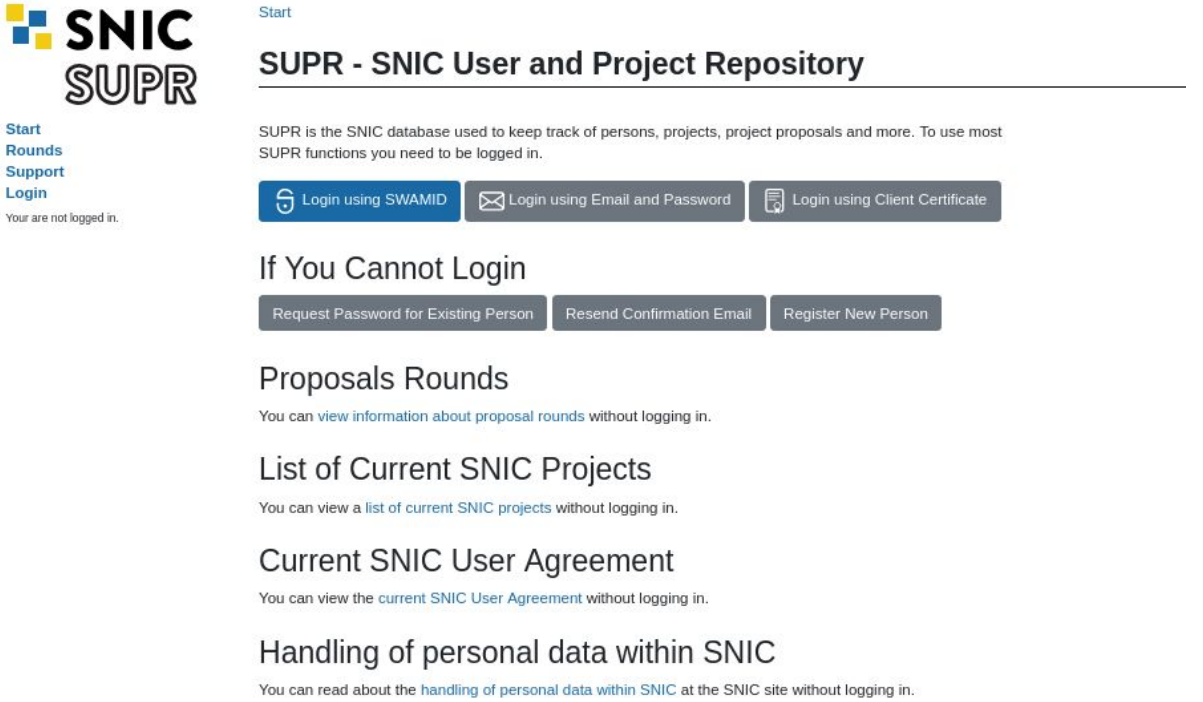
To create a new proposal, select the appropriate open round in the table above and follow the instructions on that page.

SNIC

Process

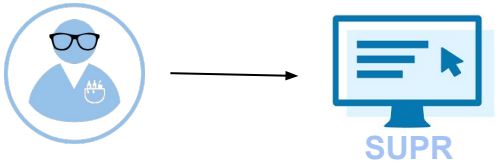


- Check for the open calls
- Register / Login to SUPR



SNIC

Process



- Check for the open calls
- Register / Login to SUPR
- Choose a round

SNIC

SNIC SUPR

Admin | User

Start
Proposals
Rounds
Projects
SNIC 2014/13-1
SNIC 2018/10-27
SNIC 2018/12-8
Groups
pdc-staff
SONC
swestore-admins
Accounts
Personal Information
Support
Logout

Logged in as:
Dejan Vitlacil
(vitlacil@kth.se)
Turn on warming colour.

Start

SUPR - Dejan Vitlacil

You are now logged in.

Proposals

You have no proposals in preparation or pending committee decision.

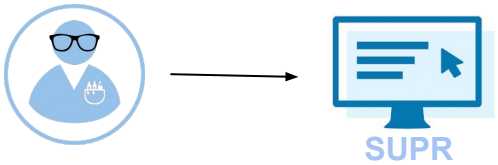
Rounds

Open for Proposals

SNIC Rounds	Deadline
SNAC Large, Spring 2019	2019-04-05 15:00
SNAC Medium, 2019	—
SNAC Small C3SE, 2019	—
SNAC Small HPC2N, 2019	—
SNAC Small Lunarc, 2019	—
SNAC Small NSC, 2019	—
SNAC Small UPPMAX, 2019	—
SNAC Science Cloud 2019	—
SNAC Medium Swestore 2019	—
SNAC Small Swestore 2019	—
SIMPLER 2019	—
SNIC SENS Medium 2019	—
SNIC SENS Small 2019	—

Local Rounds	Deadline
Cross-Border Resource sharing test instance	—
LiU Local, 2019	—
L-SENS, 2019	—
LU Local, 2019	—
LVIS, 2019	—
PDC Business 2019	—
PDC Local 2019	—
UPPMAX courses 2019	—
UPPMAX Storage 2019	—

Process



- Check for the open calls
- Register / Login to SUPR
- Choose a round
- Create and submit proposal

SNIC SUPR

Admin | User

- Start
- Proposals
- Rounds
- Projects
  - SNIC 2014/13-1
  - SNIC 2018/10-27
  - SNIC 2018/12-8
- Groups
  - pdc-staff
  - SONC
  - swestore-admins
- Accounts
- Personal Information
- Support
- Logout

Logged in as:
Dejan Vitlacil
(vitlacil@kth.se)
Turn on warning colour.

Start / Rounds / SNAC Medium Swestore 2019 / ▻ Admin

SNAC Medium Swestore 2019

This Round is Open for Proposals

More information about this round is available at https://www.snic.se/allocations/swestore/.

This round is open for proposals until 2020-01-01 00:00.

Create New Proposal for SNAC Medium Swestore 2019 | View Committee Overview

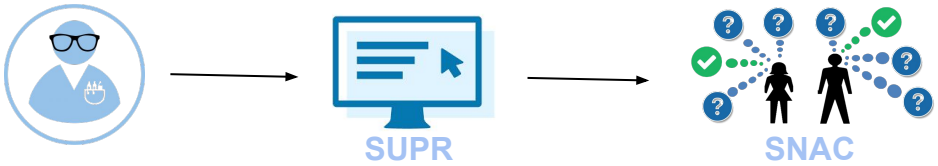
Resources

Resource	Centre	Available	Capacity Unit	Note
dCache	Swestore	1 000	TiB	*In production*
iRODS	Swestore	1 000	TiB	*In production*

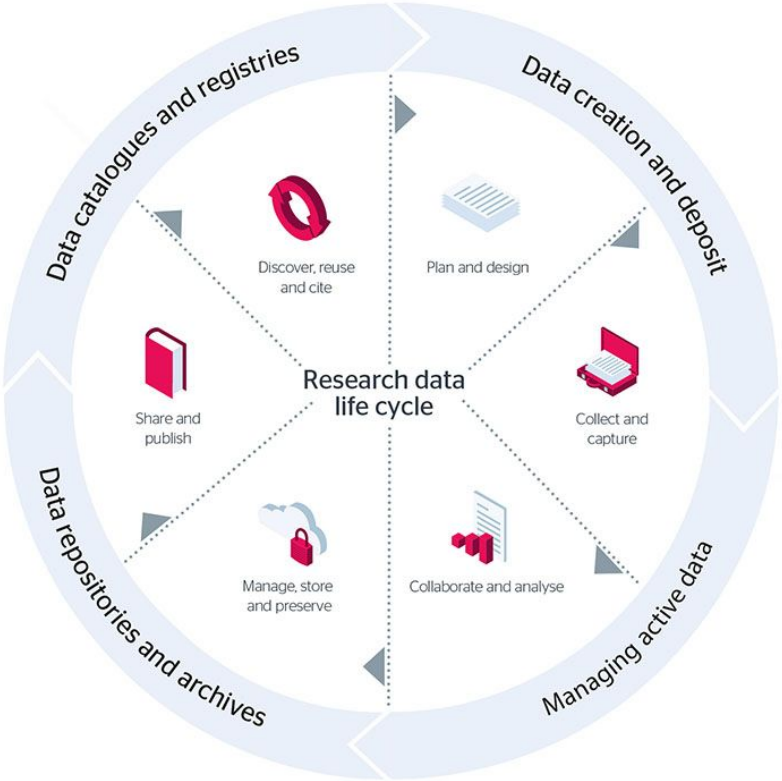
Click the ▶ to show more information about the resource.

SNIC

Process

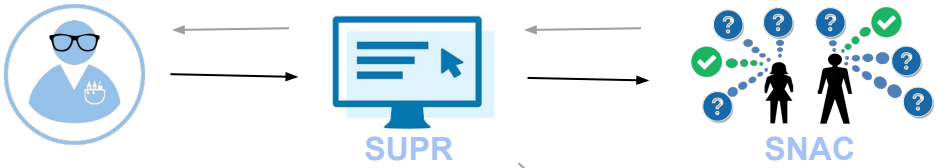


- Check for the open calls
- Register / Login to SUPR
- Choose a round
- Create and submit proposal
- DMP
- Submit to SNAC (WG)



Research data lifecycle diagram ©Jisc CC BY-NC-ND

Process



- Check for the open calls
- Register / Login to SUPR
- Choose a round
- Create and submit proposal
- DMP
- Submit to SNAC (WG)
- Proposal approved
- User is informed
- Identity Management updated

freeIPA
identity | policy | audit

- Grid Certificate
- Password (KRB/PAM)

- 14d for S and M allocation
- L allocation process missing

- Proposal approved
- Confirmation mail about proposal sent
- Change monitoring scripts
  - User
  - SNIC username
  - DN
  - Project
  - Directory (path)
  - E-mail
  - …
- Swestore information mail
  - Username
  - Project path
  - Temp password
  - Reset password
  - Documentation link
  - ....

SNIC

“Good” to know

*** PI / Researcher (person, institution or organization)**
*** Project Description**
*** Project Data Contact**
*** Related Policies**
*** .... “Administrative data” section of SND Checklist**
*****

... Much of this we already collect in SUPR - SNIC User and Project Repository!
... maybe this information could be fetchd instead from DMP tool or just referenced (PID?)

... Data Center is better then “laptop” and “USB pen”, but ...

SNIC

Not only administrative part

Fire at Samsung Group Data Center in South Korea causes service outage

Jayaram Pawar Apr 21, 2014 Innovation 1

A fire at Samsung's backup data center at Gwacheon, South Korea has knocked off services in Samsung SmartHub (Smart TV), Smartphones, Blu-Ray players and every other gadget that utilizes the servers in this particular data center for few hours before being restored. The Samsung App store and any apps that needed communication

SNIC

Common Contaminants and …

Ferrous metal particles – Ferrous metal particles come from printers, tape drives, worn parts in HVAC units new floor tiles, conduit pipes and various mechanical parts. Because these metal particles are conductive, they can cause electrical damage to circuits boards, resulting in downtime.

Concrete dust – Unsealed concrete releases fine dust into the air that consists of calcium, silica and other by-products that are abrasive and corrosive. Even when concrete is sealed, these seals can break down over time and should be inspected regularly.

Corrosive gases – Gaseous contaminants such as sulfur dioxide, hydrogen sulfide, nitrogen dioxide, chlorine, ozone and others cause copper and silver corrosion. Over time, these contaminants cause corrosion of delicate metal parts and wiring in modern circuit boards and chips.

Zinc and tin whiskers – Yes, whiskers. Whiskers are small metal filaments that grow from electroplated steel floor plates, bars in the ceiling and other metal components. Because tin and zinc are efficient conductors of electricity, these filaments can cause shorts when they break off, circulate in the air and come in contact with circuits.

Chlorides/salts – Anyone who has been to the beach knows how quickly the salty air can cause corrosion on vehicles and other metal objects. Data centers that are located in coastal areas must be careful of airborne salinity levels.

Electrostatic dust– Hazardous dust as well as dust and lint particles from clothing, cardboard, paper and other seemingly benign particles can become static-charged and interfere with servers causing data loss, erroneous commands, resets and other issues.

SNIC

Suorce: Data Foundry https://www.datafoundry.com/blog/data-center-contamination-control

Silent corruption

* data is changed unintentionally without any errors/warnings!

* cosmic rays, these neutrons occasionally cause a bit to flip from a 0 to a 1, or vice versa.

Corruption Sources

- hardware errors (memory, CPU, disk, NIC)
- data transfer noise (UTP, SATA, FC, wireless)
- firmware bugs (RAID controller, disk, NIC)
- software bugs
  - kernel (VM, filesystem, SCSI, block), libc
- Other sources (not discussed):
  - human error (be careful **and** do backups)
  - application errors/crashes (use checkpoints)
  - OS errors/crashes (rely on filesystem recovery)

SNIC

Human Error (System - Security - Management - Procurement)

1177-LÄCKAN SÄKERHET 2019-02-18 13:39

2,7 miljoner inspelade samtal till 1177 Vårdguiden helt oskyddade på internet

Lars Dobos
REPORTER

Computer Sweden avslöjar: alla telefonsamtal som ringts till 1177 sedan 2013 och som tagits emot av vårdentreprenören Medicall har legat helt öppet som ljudfiler på en oskyddad webbserver.

SNIC

“Good” to know

* PI / Researcher (person, institution or organization)
* Project Description
* Project Data Contact
* Related Policies
* .... “Administrative data” section of SND Checklist

*** Storage and backup**

What physical resources and facilities will be used for the preservation and storage of the research data? Describe how and where data is stored as well as the backup procedures for the data (retention policy). How will the data be recovered in the event of an incident? Data availability and data sharing during the project lifetime (2-3 copies geo-distributed).

*** Data protection or security policy at university (e.g. data encryption at rest or in transit)**

*** Data integrity (checksums)**

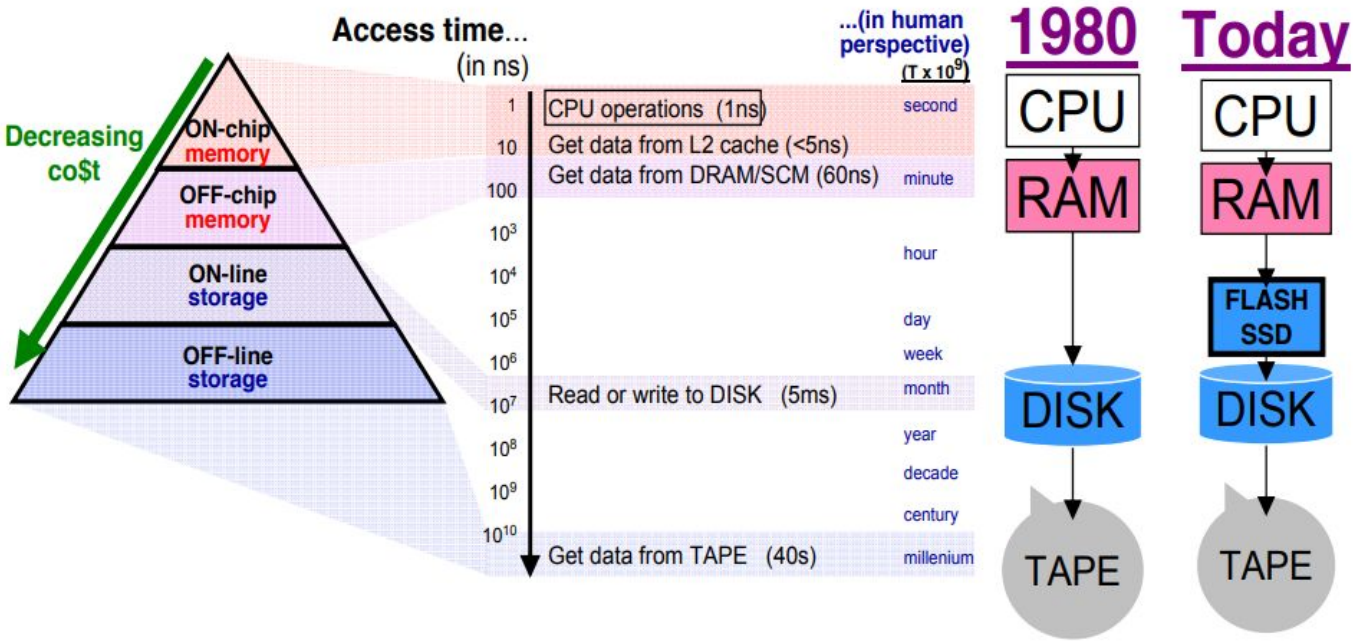
… and much more of this if there are specific needs!

… but this should also be anchored with Research Data Policy and Information Classification at University.

… do you have a well defined requirements for specific type of data?

SNIC

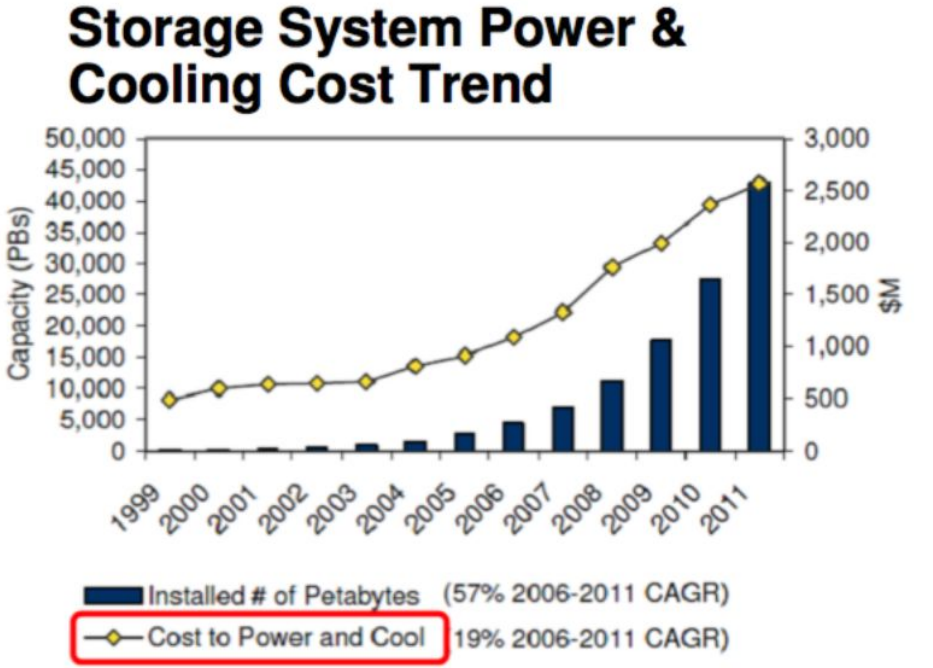
Storage class access time



picture adapted from: "Storage class memory", IBM Almaden research centre, 2013

Storage class access time

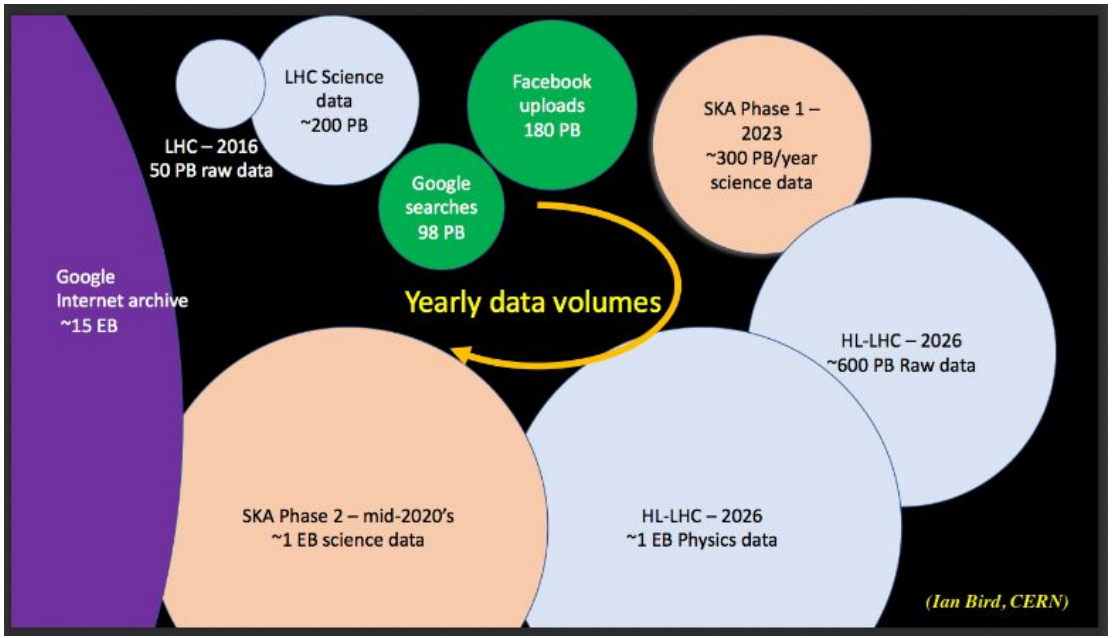
Storage systems account often for 40% of power consumption



SNIA IDC June 2008 – 'The Real Costs to Power and Cool the world's external storage'

SNIC

Predictions and growth



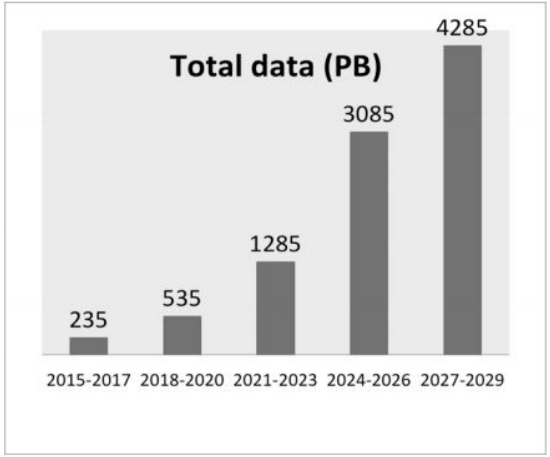
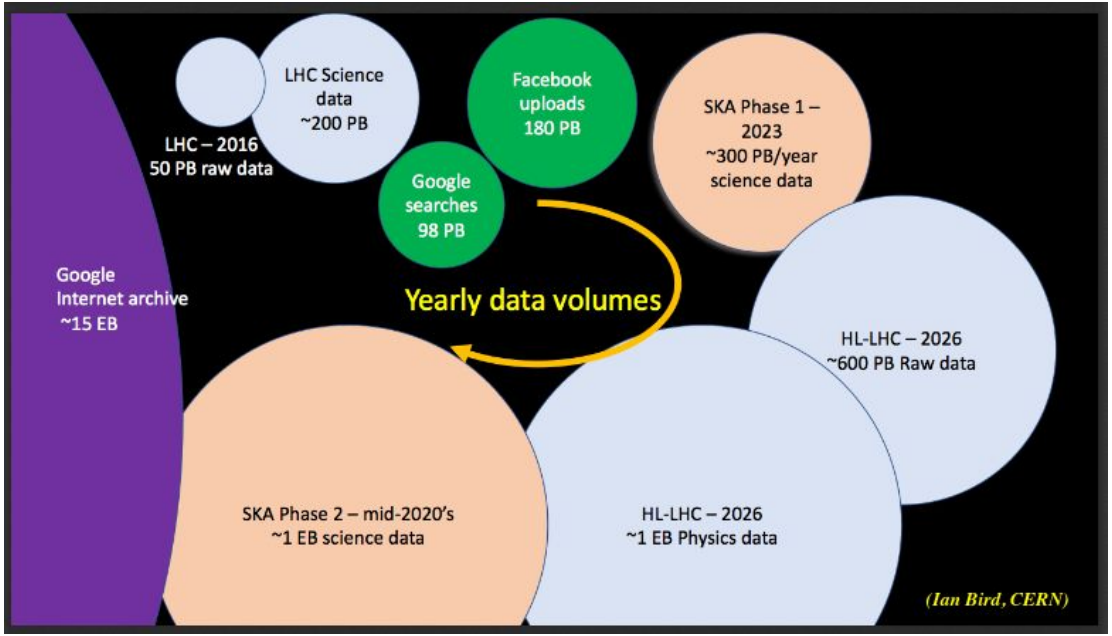


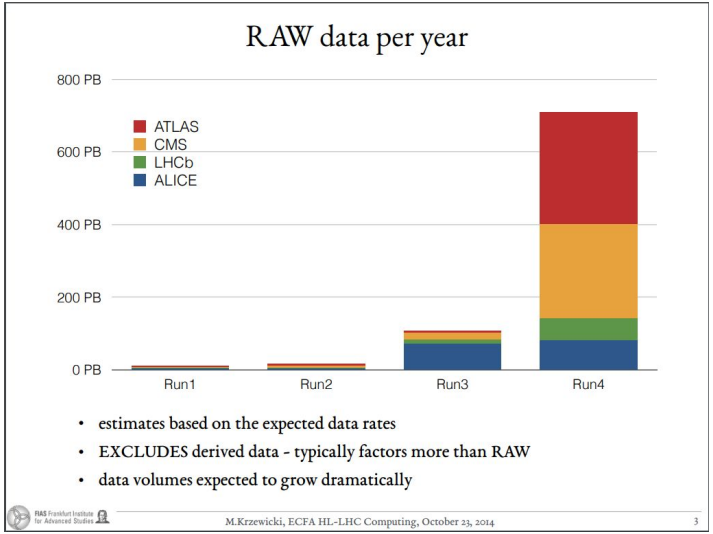
Figure 4. Predicted data growth at CERN, 2015-2029

SNIC

Journal of Physics: Conference Series 664 (2015) 042006 doi:10.1088/1742-6596/664/4/042006
Experiences and challenges running CERN's high capacity tape archive

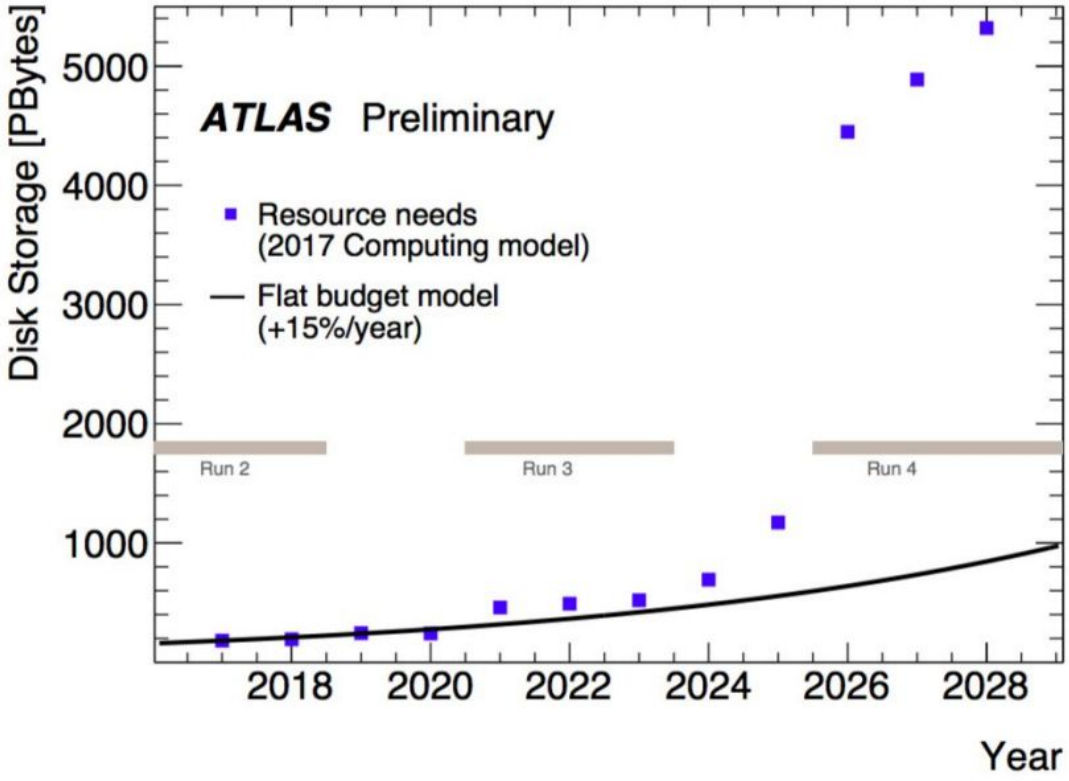
Predictions and growth





SNIC

Predictions and growth



SNIC

Research Infrastructures – Where are we going?

Research Infrastructure trends & challenges

- Internationalisation
- Diversification
- Increasingly relying on ICT
- Data deluge
- Complexity
- Trust, Authenticity
- Citation, Credits
- Open Access
- Open Data

ESFRI

European RIs:

- Around 500
- € 100 billion investment

middle age | 19th century | 20th century | 21st century

SNIC

“Good” to know

*** PI / Researcher (person, institution or organization)**
*** Project Description**
*** Project Data Contact**
*** Related Policies**
*** .... “Administrative data” section of SND Checklist**
*** Storage and backup**
What physical resources and facilities will be used for the preservation and storage of the research data? Describe how and where data is stored as well as the backup procedures for the data (retention policy). How will the data be recovered in the event of an incident? Data availability and data sharing during the project lifetime (2-3 copies geo-distributed).
*** Data protection or security policy at university (e.g. data encryption at rest or in transit)**
*** Data integrity (checksums)**
*** Use of existing data, Data collection (allocation usage roadmap, per year)**
*****

… It is important to know at which rate and when the data is coming!

SNIC

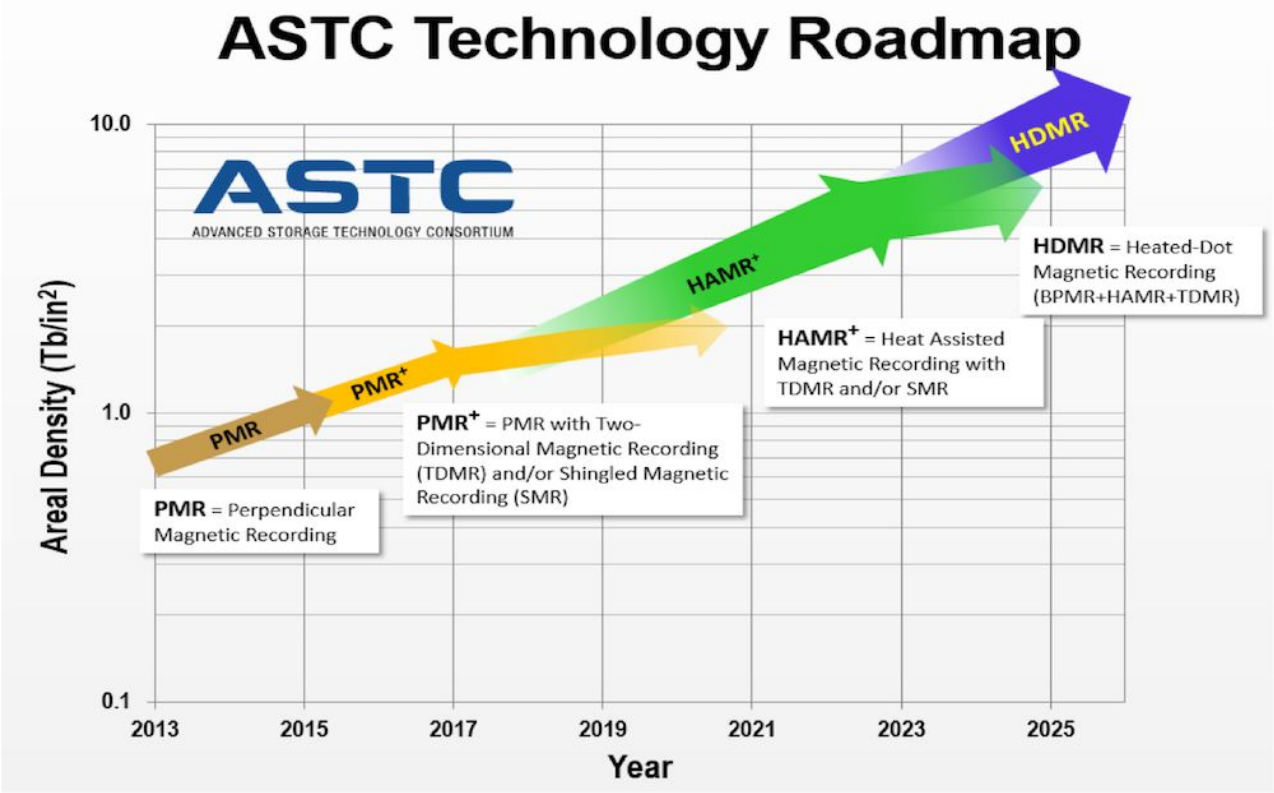
And it is not that bad, it can get worse ...

- The rain started in August and by mid-October 2011, violent floods in Thailand had crippled the factories that helped produce nearly half of the world's hard drives.
- The impact of the lost capacity was immediate as hard drive prices nearly tripled overnight.

- And how about SEK vs USD?

SNIC

And it is not that bad, it can get worse …



SNIC

“Good” to know

* PI / Researcher (person, institution or organization)
* Project Description
* Project Data Contact
* Related Policies
* .... “Administrative data” section of SND Checklist
* Storage and backup

What physical resources and facilities will be used for the preservation and storage of the research data? Describe how and where data is stored as well as the backup procedures for the data (retention policy). How will the data be recovered in the event of an incident? Data availability and data sharing during the project lifetime (2-3 copies geo-distributed).

* Data protection or security policy at university (e.g. data encryption at rest or in transit)
* Use of existing data, Data collection (allocation usage roadmap, per year)

*** Archiving/Long term preservation - (how and where should project data be delivered for preservation; if it should stay on SNIC resources, who should get access and how it should be preserved)**

Public records are documents that are either created at or has arrived to a university and shall in most cases be preserved. As long as there is not a legal support for the destruction of a document the basic principle is that it is forbidden to destroy them. Are there any routines at the university regarding public records that has to be followed? Who in the research group is responsible for making sure that the public records from the project are archived?

SNIC

EOSC Declaration

- RECOGNISING the challenges of data driven research in pursuing excellent science;
  GRANTING that the vision of European Open Science is that of a research data commons, widely inclusive of all disciplines and Member States, sustainable in the long-term, ...

- [Data Management Plans]

  A key element of good data management is a Data Management Plan (DMP); the use of DMPs should become obligatory in all research projects generating or collecting publicly funded research data, based on online tools conforming to common methodologies. ***Funder and institutional requirements must be aligned and minimum conditions for DMPs must be defined. Researchers' host institutions have a responsibility to oversee and complete the DMPs and hand them over to data repositories.***

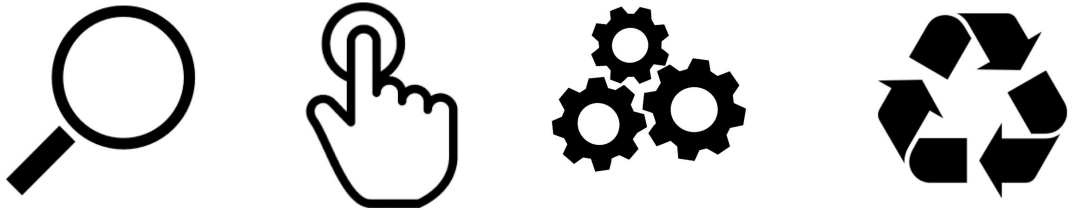
SNIC

Services

EUROPEAN OPEN SCIENCE CLOUD

https://www.eosc-portal.eu

Findable Accessible Interoperable Reusable



SNIC

International collaborations and development

EGI - European Grid Infrastructure On behalf of Vetenskapsrådet, SNIC is the Swedish partner in the EGI.eu collaboration. SNIC is partner in EC-funded project EGI-Engage. The contributions from SNIC are in the Competence Centres (CC) for BBMRI and EISCAT-3D. Read more about EGI and EGI-Engage at www.egi.eu	
EUDAT2020 is a European project funded in the H2020 programme. The project started March 2015 with a duration of three years. The project has 33 partners and a total budget of nearly 20 million Euro. Third Party for SNIC in EUDAT2020 is KTH. Home page for EUDAT.	
The PRACE infrastructure is open to European researchers. Scientists and researchers from around the world can apply for access to PRACE resources. On behalf of Vetenskapsrådet, SNIC is the Swedish partner in this collaboration. For more information about getting access to PRACE resources, click here. Visit the PRACE website for further information: http://www.prace-ri.eu/	

Worldwide LHC Computing Grid SNIC implements the Swedish contribution to the Nordic Tier-1 (NDGF) and the Swedish Tier-2 for ATLAS	
NeIC - Nordic e-Infrastructure Collaboration NeIC Home page	
EOSC-hub https://eosc-hub.eu/	
ePIC Consortium https://www.pidconsortium.eu/	
iRODS https://irods.org/	

SNIC

Questions ?

SNIC